

OPTIMAL QUANTIZERS FOR PROBABILITY DISTRIBUTIONS ON SIERPIŃSKI CARPETS

DOĞAN ÇÖMEZ AND MRINAL KANTI ROYCHOWDHURY

ABSTRACT. In this paper, we investigate the optimal sets of n -means and the n th quantization error for singular continuous probability measures on \mathbb{R}^2 supported by Sierpiński carpets. Utilizing the geometric structure of Sierpiński carpets, first we determine the optimal sets of n -means for Borel probability measures P on \mathbb{R}^2 supported by Sierpiński carpets in progressive steps. Then, using suitable Voronoi partitions associated, we obtain the number of optimal sets of n -means as well as the n th quantization error involved in each case.

1. INTRODUCTION

The history of the theory and practice of quantization dates back to 1948. Since then quantization has become an important field in electrical engineering in connection with signal processing and data compression. Broadly speaking, quantization consists in replacing an actual large data set of size N by a smaller set of prototypes of size $n \leq N$. The best choice is when loss of information about the initial data set is minimum. Quantization for probability distributions refers to the idea of estimating a given probability measure by a discrete probability measure with finite support. A good survey about the historical development of the theory has been provided by Gray and Neuhoff in [GN]. For more applied aspects of quantization the reader is referred to the book of Gersho and Gray (see [GG]). For mathematical treatment of quantization one may consult Graf-Luschgy's book (see [GL1]). Interested readers can also see [AW, GKL, GL, Z].

Let \mathbb{R}^d denote the d -dimensional Euclidean space, $\|\cdot\|$ denote the Euclidean norm on \mathbb{R}^d for any $d \geq 1$, and $n \in \mathbb{N}$. Then the n th *quantization error* for a Borel probability measure P on \mathbb{R}^d is defined by

$$V_n := V_n(P) = \inf \left\{ \int \min_{a \in \alpha} \|x - a\|^2 dP(x) : \alpha \subset \mathbb{R}^d, \text{card}(\alpha) \leq n \right\},$$

where the infimum is taken over all subsets α of \mathbb{R}^d with $\text{card}(\alpha) \leq n$. If $\int \|x\|^2 dP(x) < \infty$ then there is some set α for which the infimum is achieved (see [GL, GL1, GKL]). Such a set α for which the infimum occurs and contains no more than n points is called an *optimal set of n -means*. If α is a finite set, in general, the error $\int \min_{a \in \alpha} \|x - a\|^2 dP(x)$ is often referred to as the *variance*, *cost*, or *distortion error* for α , and is denoted by $V(P; \alpha)$. Thus, $V_n := V_n(P) = \inf \{V(P; \alpha) : \alpha \subset \mathbb{R}^d, \text{card}(\alpha) \leq n\}$. It is known that for a continuous probability measure P an optimal set of n -means always has exactly n elements (see [GL1]). Given a finite subset $\alpha \subset \mathbb{R}^d$, the Voronoi region generated by $a \in \alpha$ is defined by

$$W(a|\alpha) = \{x \in \mathbb{R}^d : \|x - a\| = \min_{b \in \alpha} \|x - b\|\}$$

i.e., the Voronoi region generated by $a \in \alpha$ is the set of all points in \mathbb{R}^d which are closest to $a \in \alpha$, and the set $\{W(a|\alpha) : a \in \alpha\}$ is called the *Voronoi diagram* or *Voronoi tessellation* of

2010 *Mathematics Subject Classification.* 60Exx, 28A80, 94A34.

Key words and phrases. Sierpiński carpet, self-affine measure, optimal quantizers, quantization error.

The research of the second author was supported by U.S. National Security Agency (NSA) Grant H98230-14-1-0320.

\mathbb{R}^d with respect to α . A Borel measurable partition $\{A_a : a \in \alpha\}$ of \mathbb{R}^d is called a *Voronoi partition* of \mathbb{R}^d with respect to α (and P) if

$$A_a \subset W(a|\alpha) \text{ (} P\text{-a.e.) for every } a \in \alpha.$$

Note that if $\alpha = \{a_1, a_2, \dots, a_n\}$ is an optimal set of n -means for P and $\{A_1, A_2, \dots, A_n\}$ is a Voronoi partition with respect to α , then

$$V_n = \sum_{i=1}^n \int_{A_i} \|x - a_i\|^2 dP(x).$$

A *centroidal Voronoi tessellation* (CVT) is a Voronoi tessellation in which the generators are the centroids for each Voronoi region. A CVT with n generators, also called a CVT with n -means, associated with a probability measure P is called an *optimal centroidal Voronoi tessellation* (OCVT) if the n generators form an optimal set of n -means for P . The following proposition provides further information on the Voronoi regions generated by an optimal set of n -means (see [GG, GL1]).

Proposition 1.1. *Let α be an optimal set of n -means, $a \in \alpha$, and $M(a)$ be the Voronoi region generated by $a \in \alpha$, i.e.,*

$$M(a) = \{x \in \mathbb{R}^d : \|x - a\| = \min_{b \in \alpha} \|x - b\|\}.$$

Then for every $a \in \alpha$,

- (i) $P(M(a)) > 0$,
- (ii) $P(\partial M(a)) = 0$,
- (iii) $a = E(X : X \in M(a))$, and
- (iv) P -almost surely the set $\{M(a) : a \in \alpha\}$ forms a Voronoi partition of \mathbb{R}^d .

Remark 1.2. Let α be an optimal set of n -means and $a \in \alpha$, then by Proposition 1.1, we have

$$a = \frac{1}{P(M(a))} \int_{M(a)} x dP = \frac{\int_{M(a)} x dP}{\int_{M(a)} dP},$$

which implies that a is the centroid of the Voronoi region $M(a)$ associated with the probability measure P (see also [DFG]). Thus, we can say that for a Borel probability measure P on \mathbb{R}^d , an optimal set of n -means forms a centroidal Voronoi tessellation of \mathbb{R}^d ; however, the converse is not true in general (see [DFG, GG]).

It is known that the classical Cantor set C is generated by the two contractive similarity mappings S_1 and S_2 given by $S_1(x) = \frac{1}{3}x$ and $S_2(x) = \frac{1}{3}x + \frac{2}{3}$ for all $x \in \mathbb{R}$. Let P be a Borel probability measure on \mathbb{R} such that $P = \frac{1}{2}P \circ S_1^{-1} + \frac{1}{2}P \circ S_2^{-1}$, where $P \circ S_i^{-1}$ denotes the image measure of P with respect to S_i for $i = 1, 2$ (see [H], Theorem 4.4(1) for a generalization of self-similar measure). Then, P has support the Cantor set C . For this probability measure Graf and Luschgy determined the optimal sets of n -means and the n th quantization error for all $n \geq 1$ (see [GL2]). In this paper, we have considered a Sierpiński carpet, denoted by S , which is generated by the four contractive similarity mappings S_1, S_2, S_3 and S_4 on \mathbb{R}^2 such that $S_1(x_1, x_2) = \frac{1}{3}(x_1, x_2)$, $S_2(x_1, x_2) = \frac{1}{3}(x_1, x_2) + (\frac{2}{3}, 0)$, $S_3(x_1, x_2) = \frac{1}{3}(x_1, x_2) + (0, \frac{2}{3})$, and $S_4(x_1, x_2) = \frac{1}{3}(x_1, x_2) + (\frac{2}{3}, \frac{2}{3})$ for all $(x_1, x_2) \in \mathbb{R}^2$. If P is a Borel probability measure on \mathbb{R}^2 such that $P = \frac{1}{4}P \circ S_1^{-1} + \frac{1}{4}P \circ S_2^{-1} + \frac{1}{4}P \circ S_3^{-1} + \frac{1}{4}P \circ S_4^{-1}$, then P has support the Sierpiński carpet S . For this probability measure P , in this paper we have determined the optimal sets of n -means and the n th quantization error. For the Cantor distribution, for any n -points, one can easily determine whether the n -points form a CVT (see [GL2]), but for the probability distribution supported by the Sierpiński carpet, considered in our paper, for

n -points sometimes it is quite difficult whether the points form a CVT. The technique we utilized can be extended to determine the optimal sets and the corresponding quantization error for many other singular continuous probability measures, such as probability measures on more general Sierpiński carpets, probability measures on Sierpiński gaskets, etc.

2. BASIC DEFINITIONS AND LEMMAS

In this section, we give the basic definitions and lemmas that will be instrumental in our analysis. By a *string* or a *word* σ over an alphabet $I = \{1, 2, 3, 4\}$, we mean a finite sequence $\sigma := \sigma_1 \sigma_2 \cdots \sigma_k$ of symbols from the alphabet, where $k \geq 1$, and k is called the length of the word σ . A word of length zero is called the *empty word*, and is denoted by \emptyset . By I^* we denote the set of all words over the alphabet I of some finite length k including the empty word \emptyset . By $|\sigma|$, we denote the length of a word $\sigma \in I^*$. For any two words $\sigma := \sigma_1 \sigma_2 \cdots \sigma_k$ and $\tau := \tau_1 \tau_2 \cdots \tau_\ell$ in I^* , by $\sigma\tau := \sigma_1 \cdots \sigma_k \tau_1 \cdots \tau_\ell$ we mean the word obtained from the concatenation of the two words σ and τ . The maps $S_i : \mathbb{R}^2 \rightarrow \mathbb{R}^2$, $1 \leq i \leq 4$, will be the generating maps of the Sierpiński carpet defined as before. For $\sigma = \sigma_1 \sigma_2 \cdots \sigma_k \in I^k$, set $S_\sigma = S_{\sigma_1} \circ \cdots \circ S_{\sigma_k}$ and $J_\sigma = S_\sigma([0, 1] \times [0, 1])$. For the empty word \emptyset , by S_\emptyset we mean the identity mapping on \mathbb{R}^2 , and write $J = J_\emptyset = S_\emptyset([0, 1] \times [0, 1]) = [0, 1] \times [0, 1]$. The sets $\{J_\sigma : \sigma \in \{1, 2, 3, 4\}^k\}$ are just the 4^k squares in the k th level in the construction of the Sierpiński carpet. The squares J_{σ_1} , J_{σ_2} , J_{σ_3} and J_{σ_4} into which J_σ is split up at the $(k+1)$ th level are called the children of J_σ . By the center of J_σ , we mean the point of intersection of the two diagonals of J_σ . The set $S = \bigcap_{k \in \mathbb{N}} \bigcup_{\sigma \in \{1, 2, 3, 4\}^k} J_\sigma$ is the Sierpiński carpet and equals the support of the probability measure P given by $P = \frac{1}{4}P \circ S_1^{-1} + \frac{1}{4}P \circ S_2^{-1} + \frac{1}{4}P \circ S_3^{-1} + \frac{1}{4}P \circ S_4^{-1}$.

Let us now give the following lemma.

Lemma 2.1. *Let $f : \mathbb{R} \rightarrow \mathbb{R}^+$ be Borel measurable and $k \in \mathbb{N}$. Then,*

$$\int f dP = \frac{1}{4^k} \sum_{\sigma \in I^k} \int f \circ S_\sigma dP.$$

Proof. We know $P = \frac{1}{4}P \circ S_1^{-1} + \frac{1}{4}P \circ S_2^{-1} + \frac{1}{4}P \circ S_3^{-1} + \frac{1}{4}P \circ S_4^{-1}$, and so by induction $P = \sum_{\sigma \in I^k} \frac{1}{4^k} P \circ S_\sigma^{-1}$, and thus the lemma is yielded. \square

Let $S_{(i1)}$, $S_{(i2)}$ be the horizontal and vertical components of the transformation S_i for $i = 1, 2, 3, 4$. Then for any $(x_1, x_2) \in \mathbb{R}^2$ we have $S_{(11)}(x_1) = \frac{1}{3}x_1$, $S_{(12)}(x_2) = \frac{1}{3}x_2$, $S_{(21)}(x_1) = \frac{1}{3}x_1 + \frac{2}{3}$, $S_{(22)}(x_2) = \frac{1}{3}x_2$, $S_{(31)}(x_1) = \frac{1}{3}x_1$, $S_{(32)}(x_2) = \frac{1}{3}x_2 + \frac{2}{3}$, and $S_{(41)}(x_1) = \frac{1}{3}x_1 + \frac{2}{3}$, $S_{(42)}(x_2) = \frac{1}{3}x_2 + \frac{2}{3}$. Let $X = (X_1, X_2)$ be a bivariate random variable with distribution P . Let P_1, P_2 be the marginal distributions of P , i.e., $P_1(A) = P(A \times \mathbb{R})$ for all $A \in \mathfrak{B}$, and $P_2(B) = P(\mathbb{R} \times B)$ for all $B \in \mathfrak{B}$. Here \mathfrak{B} is the Borel σ -algebra on \mathbb{R} . Then X_1 has distribution P_1 and X_2 has distribution P_2 .

The statement below provides the connection between P and its marginal distributions via the components of the generating maps S_i .

Lemma 2.2. *Let P_1 and P_2 be the marginal distributions of the probability measure P . Then,*

$$\begin{aligned} P_1 &= \frac{1}{4}P_1 \circ S_{(11)}^{-1} + \frac{1}{4}P_1 \circ S_{(21)}^{-1} + \frac{1}{4}P_1 \circ S_{(31)}^{-1} + \frac{1}{4}P_1 \circ S_{(41)}^{-1} \text{ and} \\ P_2 &= \frac{1}{4}P_1 \circ S_{(12)}^{-1} + \frac{1}{4}P_1 \circ S_{(22)}^{-1} + \frac{1}{4}P_1 \circ S_{(32)}^{-1} + \frac{1}{4}P_1 \circ S_{(42)}^{-1}. \end{aligned}$$

Proof. Let us take any $A \in \mathfrak{B}(\mathbb{R})$. Then

$$\begin{aligned}
& \left(\frac{1}{4}P_1 \circ S_{(11)}^{-1} + \frac{1}{4}P_1 \circ S_{(21)}^{-1} + \frac{1}{4}P_1 \circ S_{(31)}^{-1} + \frac{1}{4}P_1 \circ S_{(41)}^{-1} \right)(A) \\
&= \frac{1}{4}P_1 \circ S_{(11)}^{-1}(A) + \frac{1}{4}P_1 \circ S_{(21)}^{-1}(A) + \frac{1}{4}P_1 \circ S_{(31)}^{-1}(A) + \frac{1}{4}P_1 \circ S_{(41)}^{-1}(A) \\
&= \frac{1}{4}P_1(3A) + \frac{1}{4}P_1(3(A - \frac{2}{3})) + \frac{1}{4}P_1(3A) + \frac{1}{4}P_1(3(A - \frac{2}{3})) \\
&= \frac{1}{4}P(3A \times \mathbb{R}) + \frac{1}{4}P(3(A - \frac{2}{3}) \times \mathbb{R}) + \frac{1}{4}P(3A \times \mathbb{R}) + \frac{1}{4}P(3(A - \frac{2}{3}) \times \mathbb{R}) \\
&= \frac{1}{4}P \circ S_1^{-1}(A \times \mathbb{R}) + \frac{1}{4}P \circ S_2^{-1}(A \times \mathbb{R}) + \frac{1}{4}P \circ S_3^{-1}(A \times \mathbb{R}) + \frac{1}{4}P \circ S_4^{-1}(A \times \mathbb{R}) \\
&= P(A \times \mathbb{R}) \\
&= P_1(A),
\end{aligned}$$

and thus the expression for P_1 follows. The expression for P_2 is proved similarly. \square

For words $\beta, \gamma, \dots, \delta$ in I^* , by $a(\beta, \gamma, \dots, \delta)$ we mean the conditional expectation of the random variable X given $J_\beta \cup J_\gamma \cup \dots \cup J_\delta$, i.e.,

$$(1) \quad a(\beta, \gamma, \dots, \delta) = E(X | X \in J_\beta \cup J_\gamma \cup \dots \cup J_\delta) = \frac{1}{P(J_\beta \cup \dots \cup J_\delta)} \int_{J_\beta \cup \dots \cup J_\delta} x dP.$$

By $\int x dP$ it is meant $\int (x_1, x_2) dP$. Let us now give the following lemma.

Lemma 2.3. *Let $E(X)$ and $V(X)$ denote the expectation and the variance of the random variable X . Then,*

$$E(X) = (E(X_1), E(X_2)) = \left(\frac{1}{2}, \frac{1}{2}\right) \text{ and } V := V(X) = E\|X - \left(\frac{1}{2}, \frac{1}{2}\right)\|^2 = \frac{1}{4}.$$

Proof. We have

$$\begin{aligned}
& E(X_1) \\
&= \int x dP_1 = \frac{1}{4} \int x dP_1 \circ S_{(11)}^{-1} + \frac{1}{4} \int x dP_1 \circ S_{(21)}^{-1} + \frac{1}{4} \int x dP_1 \circ S_{(31)}^{-1} + \frac{1}{4} \int x dP_1 \circ S_{(41)}^{-1} \\
&= \frac{1}{4} \int \frac{1}{3}x dP_1 + \frac{1}{4} \int \left(\frac{1}{3}x + \frac{2}{3}\right) dP_1 + \frac{1}{4} \int \frac{1}{3}x dP_1 + \frac{1}{4} \int \left(\frac{1}{3}x + \frac{2}{3}\right) dP_1 \\
&= \frac{1}{12}E(X_1) + \frac{1}{12}E(X_1) + \frac{2}{12} + \frac{1}{12}E(X_1) + \frac{1}{12}E(X_1) + \frac{2}{12} \\
&= \frac{1}{3}E(X_1) + \frac{1}{3},
\end{aligned}$$

and thus $E(X_1) = \frac{1}{2}$. Similarly one can show that $E(X_2) = \frac{1}{2}$. Now,

$$\begin{aligned}
E(X_1^2) &= \int x^2 dP_1 \\
&= \frac{1}{4} \int x^2 dP_1 \circ S_{(11)}^{-1} + \frac{1}{4} \int x^2 dP_1 \circ S_{(21)}^{-1} + \frac{1}{4} \int x^2 dP_1 \circ S_{(31)}^{-1} + \frac{1}{4} \int x^2 dP_1 \circ S_{(41)}^{-1} \\
&= \frac{1}{4} \int \left(\frac{1}{3}x\right)^2 dP_1 + \frac{1}{4} \int \left(\frac{1}{3}x + \frac{2}{3}\right)^2 dP_1 + \frac{1}{4} \int \left(\frac{1}{3}x\right)^2 dP_1 + \frac{1}{4} \int \left(\frac{1}{3}x + \frac{2}{3}\right)^2 dP_1 \\
&= 2 \left(\frac{1}{4} \int \frac{1}{9} x^2 dP_1 + \frac{1}{4} \int \left(\frac{1}{9} x^2 + \frac{4}{9} x + \frac{4}{9}\right) dP_1 \right) \\
&= 2 \left(\frac{1}{36} E(X_1^2) + \frac{1}{36} E(X_1^2) + \frac{1}{9} E(X_1) + \frac{1}{9} \right) \\
&= \frac{1}{9} E(X_1^2) + \frac{2}{9} E(X_1) + \frac{2}{9}.
\end{aligned}$$

This implies $E(X_1^2) = \frac{3}{8}$. Thus, $V(X_1) = E(X_1^2) - (E(X_1))^2 = \frac{3}{8} - \frac{1}{4} = \frac{1}{8}$, and similarly $V(X_2) = \frac{1}{8}$. Hence,

$$\begin{aligned}
E\|X - \left(\frac{1}{2}, \frac{1}{2}\right)\|^2 &= \iint_{\mathbb{R}^2} \left[(x_1 - \frac{1}{2})^2 + (x_2 - \frac{1}{2})^2\right] dP(x_1, x_2) \\
&= \int_{x_1=-\infty}^{\infty} (x_1 - \frac{1}{2})^2 \int_{x_2=-\infty}^{\infty} dP(x_1, x_2) + \int_{x_2=-\infty}^{\infty} (x_2 - \frac{1}{2})^2 \int_{x_1=-\infty}^{\infty} dP(x_1, x_2) \\
&= \int_0^1 (x_1 - \frac{1}{2})^2 dP_1(x_1) + \int_0^1 (x_2 - \frac{1}{2})^2 dP_2(x_2) \\
&= E(X_1 - \frac{1}{2})^2 + E(X_2 - \frac{1}{2})^2 \\
&= V(X_1) + V(X_2) = \frac{1}{4}.
\end{aligned}$$

Hence the lemma follows. \square

Now, the following two notes are in order.

Note 2.4. We have $E(X_1) = \frac{1}{2}$ and $E(X_2) = \frac{1}{2}$, and so by the standard rule of probability theory, for any two real numbers a and b ,

$$\begin{aligned}
E(X_1 - a)^2 &= E(X_1 - \frac{1}{2})^2 + (a - \frac{1}{2})^2 = V(X_1) + (a - \frac{1}{2})^2, \text{ and similarly} \\
E(X_2 - b)^2 &= V(X_2) + (b - \frac{1}{2})^2.
\end{aligned}$$

Thus, for any $(a, b) \in \mathbb{R}^2$,

$$\begin{aligned}
E\|X - (a, b)\|^2 &= \iint_{\mathbb{R}^2} [(x_1 - a)^2 + (x_2 - b)^2] dP(x_1, x_2) \\
&= \int_{x_1=-\infty}^{\infty} (x_1 - a)^2 \int_{x_2=-\infty}^{\infty} dP(x_1, x_2) + \int_{x_2=-\infty}^{\infty} (x_2 - b)^2 \int_{x_1=-\infty}^{\infty} dP(x_1, x_2) \\
&= \int_0^1 (x_1 - a)^2 dP_1(x_1) + \int_0^1 (x_2 - b)^2 dP_2(x_2) = E(X_1 - a)^2 + E(X_2 - b)^2 \\
&= V(X_1) + V(X_2) + (a - \frac{1}{2})^2 + (b - \frac{1}{2})^2 = V + \|(a, b) - (\frac{1}{2}, \frac{1}{2})\|^2.
\end{aligned}$$

In fact, for any $\sigma \in I^k$, $k \geq 1$, we have

$$\int_{J_\sigma} \|x - (a, b)\|^2 dP = \frac{1}{4^k} \int \|(x_1, x_2) - (a, b)\|^2 dP \circ S_\sigma^{-1},$$

which implies

$$(2) \quad \int_{J_\sigma} \|x - (a, b)\|^2 dP = \frac{1}{4^k} \left(\frac{1}{9^k} V + \|S_\sigma(\frac{1}{2}, \frac{1}{2}) - (a, b)\|^2 \right).$$

Note 2.5. From Lemma 2.3 it follows that the optimal set of one-mean is the expected value and the corresponding quantization error is the variance V of the random variable X . For $\sigma \in I^k$, $k \geq 1$, since $a(\sigma) = E(X : X \in J_\sigma)$, using Lemma 2.1, we have

$$a(\sigma) = \frac{1}{P(J_\sigma)} \int_{J_\sigma} x dP(x) = \int_{J_\sigma} x dP \circ S_\sigma^{-1}(x) = \int S_\sigma(x) dP(x) = E(S_\sigma(X)).$$

Since S_i is a similarity mapping, it is easy to see that $E(S_j(X)) = S_j(E(X))$ for $j = 1, 2, 3, 4$ and so by induction, $a(\sigma) = E(S_\sigma(X)) = S_\sigma(E(X)) = S_\sigma(\frac{1}{2}, \frac{1}{2})$ for $\sigma \in I^k$, $k \geq 1$.

In the next two sections, we will determine the optimal sets of two-, three-, and four-means. They are the key to understand the configuration of optimal sets of n -means for all $n \geq 5$. Then, in Section 5, we will analyze the optimal sets of n -means for all $n \geq 5$.

3. OPTIMAL SETS OF 2-MEANS

In this section we obtain all the optimal sets of two-means and the corresponding quantization error.

Let J be the unit square with vertices $O(0, 0)$, $A(1, 0)$, $B(1, 1)$ and $C(0, 1)$. Recall that an optimal set of two-means forms a CVT with two-means. If possible, let us first assume that the two points in an optimal set of two-means lie on the vertical line ℓ which is a perpendicular bisector of the two opposite sides of the square. Then, the boundary of the two Voronoi regions will be the horizontal line m which is a perpendicular bisector of the other two opposite sides. This is true since the children generating the Sierpiński carpet at each level have equal weight with respect to the probability measure P and are symmetrically distributed over the square. Thus if $P(p_1, p_2)$ and $Q(q_1, q_2)$ are the two points in this case, then

$$\begin{aligned} (p_1, p_2) &= E(X : X \in J_3 \cup J_4) = \frac{1}{P(J_3 \cup J_4)} \int_{J_3 \cup J_4} (x_1, x_2) dP \\ &= 2 \left(\int_{J_3} (x_1, x_2) dP + \int_{J_4} (x_1, x_2) dP \right) = 2 \left(\frac{1}{4} \int (x_1, x_2) dP \circ S_3^{-1} + \frac{1}{4} \int (x_1, x_2) dP \circ S_4^{-1} \right) \\ &= \frac{1}{2} \left(S_3\left(\frac{1}{2}, \frac{1}{2}\right) + S_4\left(\frac{1}{2}, \frac{1}{2}\right) \right) = \left(\frac{1}{2}, \frac{5}{6}\right). \end{aligned}$$

Similarly one can show that $(q_1, q_2) = E(X : X \in J_1 \cup J_2) = (\frac{1}{2}, \frac{1}{6})$. If $V_{2,1}$ is the distortion error in this case, then using (2), we have

$$\begin{aligned} V_{2,1} &= \int_{J_1 \cup J_2} \|(x_1, x_2) - (\frac{1}{2}, \frac{1}{6})\|^2 dP + \int_{J_3 \cup J_4} \|(x_1, x_2) - (\frac{1}{2}, \frac{5}{6})\|^2 dP \\ &= \frac{1}{4} \left(\frac{4}{9} V + \|S_1(\frac{1}{2}, \frac{1}{2}) - (\frac{1}{2}, \frac{1}{6})\|^2 + \|S_2(\frac{1}{2}, \frac{1}{2}) - (\frac{1}{2}, \frac{1}{6})\|^2 + \|S_3(\frac{1}{2}, \frac{1}{2}) - (\frac{1}{2}, \frac{5}{6})\|^2 \right. \\ &\quad \left. + \|S_4(\frac{1}{2}, \frac{1}{2}) - (\frac{1}{2}, \frac{5}{6})\|^2 \right) \\ &= \frac{1}{4} \left(\frac{4}{9} V + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} \right) = \frac{5}{36} = 0.138889. \end{aligned}$$

Due to symmetry of the lines ℓ and m , if we assume that the two points in an optimal set of two-means are $(\frac{1}{6}, \frac{1}{2})$ and $(\frac{5}{6}, \frac{1}{2})$, the distortion error will also be equal to $V_{2,1}$. If possible, let us now assume that the two points in an optimal set of two-means (p_1, p_2) and (q_1, q_2) lie on a diagonal, say AC , of the square, and then the other diagonal OB is the boundary of the Voronoi regions in this case. We now look at the following two sums:

$$\begin{aligned} & \frac{5}{24} + \frac{11}{144} + \frac{29}{864} + \frac{83}{5184} + \frac{245}{31104} + \frac{731}{186624} + \frac{2189}{1119744} + \dots \\ &= \frac{1}{4} \left(\frac{5}{6} + \frac{11}{6^2} + \frac{29}{6^3} + \frac{29}{6^4} + \frac{83}{6^5} + \frac{245}{6^6} + \frac{731}{6^7} + \dots \right) \\ &= \frac{1}{4} \sum_{i=1}^{\infty} \frac{3^i + 2}{6^i} = \frac{1}{4} \sum_{i=1}^{\infty} \frac{1}{2^i} + \frac{1}{2} \sum_{i=1}^{\infty} \frac{1}{6^i} = \frac{1}{4} \cdot \frac{\frac{1}{2}}{1 - \frac{1}{2}} + \frac{1}{2} \cdot \frac{\frac{1}{6}}{1 - \frac{1}{6}} = \frac{7}{20}, \end{aligned}$$

and

$$\begin{aligned} & \frac{1}{24} + \frac{7}{144} + \frac{25}{864} + \frac{79}{5184} + \frac{241}{31104} + \frac{727}{186624} + \frac{2185}{1119744} + \dots \\ &= \frac{1}{4} \left(\frac{1}{6} + \frac{7}{6^2} + \frac{25}{6^3} + \frac{79}{6^4} + \frac{241}{6^5} + \frac{727}{6^6} + \frac{2185}{6^7} + \dots \right) \\ &= \frac{1}{4} \sum_{i=1}^{\infty} \frac{3^i - 2}{6^i} = \frac{1}{4} \sum_{i=1}^{\infty} \frac{1}{2^i} - \frac{1}{2} \sum_{i=1}^{\infty} \frac{1}{6^i} = \frac{1}{4} \cdot \frac{\frac{1}{2}}{1 - \frac{1}{2}} - \frac{1}{2} \cdot \frac{\frac{1}{6}}{1 - \frac{1}{6}} = \frac{3}{20}. \end{aligned}$$

Thus, using (1), we have

$$\begin{aligned} (3) \quad & (p_1, p_2) = E(X : X \in J_2 \cup (J_{12} \cup J_{42}) \cup (J_{112} \cup J_{142} \cup J_{412} \cup J_{442}) \cup \dots) \\ &= 2 \left(\frac{1}{4} S_2\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{4^2} \left(S_{12}\left(\frac{1}{2}, \frac{1}{2}\right) + S_{42}\left(\frac{1}{2}, \frac{1}{2}\right) \right) + \frac{1}{4^3} \left(S_{112}\left(\frac{1}{2}, \frac{1}{2}\right) + S_{142}\left(\frac{1}{2}, \frac{1}{2}\right) \right. \right. \\ & \quad \left. \left. + S_{412}\left(\frac{1}{2}, \frac{1}{2}\right) + S_{442}\left(\frac{1}{2}, \frac{1}{2}\right) \right) + \dots \right) \\ &= 2 \left(\left(\frac{5}{24}, \frac{1}{24} \right) + \left(\frac{11}{144}, \frac{7}{144} \right) + \left(\frac{29}{864}, \frac{25}{864} \right) + \left(\frac{83}{5184}, \frac{79}{5184} \right) + \dots \right) \\ &= 2 \left(\frac{5}{24} + \frac{11}{144} + \frac{29}{864} + \frac{83}{5184} + \dots, \frac{1}{24} + \frac{7}{144} + \frac{25}{864} + \frac{79}{5184} + \dots \right) \\ &= 2 \left(\frac{7}{20}, \frac{3}{20} \right) = \left(\frac{7}{10}, \frac{3}{10} \right). \end{aligned}$$

Similarly, one can show that

$$(4) \quad (q_1, q_2) = E(X : X \in J_3 \cup (J_{13} \cup J_{43}) \cup (J_{113} \cup J_{143} \cup J_{413} \cup J_{443}) \cup \dots) = \left(\frac{3}{10}, \frac{7}{10} \right).$$

Hence, if $V_{2,2}$ is the distortion error due to the points $(\frac{7}{10}, \frac{3}{10})$ and $(\frac{3}{10}, \frac{7}{10})$ in this case, using (2), we have

$$\begin{aligned} V_{2,2} &= 2 \left(\text{distortion error due to the point } \left(\frac{7}{10}, \frac{3}{10} \right) \right) \\ &> 2 \left(\int_{J_2} \|(x_1, x_2) - \left(\frac{7}{10}, \frac{3}{10} \right)\|^2 dP + \int_{J_{12} \cup J_{42}} \|(x_1, x_2) - \left(\frac{7}{10}, \frac{3}{10} \right)\|^2 dP \right. \\ & \quad \left. + \int_{J_{112} \cup J_{412} \cup J_{142} \cup J_{442}} \|(x_1, x_2) - \left(\frac{7}{10}, \frac{3}{10} \right)\|^2 dP \right. \\ & \quad \left. + \int_{J_{1112} \cup J_{4112} \cup J_{1142} \cup J_{4142} \cup J_{4112} \cup J_{4412} \cup J_{4142} \cup J_{4442}} \|(x_1, x_2) - \left(\frac{7}{10}, \frac{3}{10} \right)\|^2 dP \right) \\ &= 2(0.0747492) = 0.1494984 > V_{2,1}. \end{aligned}$$

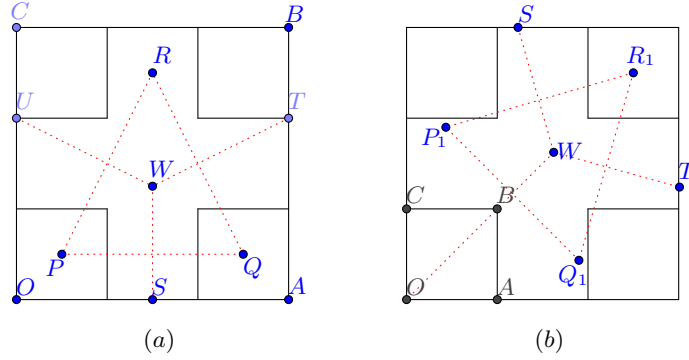


FIGURE 1. (a) CVT with three-means P , Q and R with one on the vertical line of the symmetry; (b) CVT with three-means P_1 , Q_1 and R_1 with one on a diagonal of the square.

Though the calculations are cumbersome, similarly one can show that if the two points in an optimal set of two-means lie on any other transversal, except ℓ and m , the distortion error will be larger than $V_{2,1}$. Thus, we deduce the following proposition.

Proposition 3.1. *Let P be a Borel probability measure on \mathbb{R}^2 supported by the Sierpiński carpet as defined before. Then the sets $\{(\frac{1}{6}, \frac{1}{2}), (\frac{5}{6}, \frac{1}{2})\}$ and $\{(\frac{1}{2}, \frac{1}{6}), (\frac{1}{2}, \frac{5}{6})\}$ form two different optimal sets of two-means with quantization error $\frac{5}{36}$.*

4. OPTIMAL SETS OF 3-MEANS

In this section we determine the optimal sets of three-means. Let us first prove the following lemma.

Lemma 4.1. *Let P be a Borel probability measure on \mathbb{R}^2 supported by the Sierpiński carpet as defined before. Then the set $\alpha_3 = \{(\frac{1}{6}, \frac{1}{6}), (\frac{5}{6}, \frac{1}{6}), (\frac{1}{2}, \frac{5}{6})\}$ forms a CVT with three-means and the corresponding distortion error is $\frac{1}{12}$.*

Proof. Recall that the boundaries of the Voronoi regions lie along the perpendicular bisectors of the line segments joining their centers. The perpendicular bisectors of the line segments joining each pair of points from the list $\{(\frac{1}{6}, \frac{1}{6}), (\frac{5}{6}, \frac{1}{6}), (\frac{1}{2}, \frac{5}{6})\}$ are SW , TW and UW with equations respectively $x_1 = \frac{1}{2}$, $x_2 = \frac{1}{2}x_1 + \frac{1}{6}$ and $x_2 = -\frac{1}{2}x_1 + \frac{2}{3}$, and they concur at the point $W(\frac{1}{2}, \frac{5}{12})$ as shown in Figure 1 (a). Thus, the three regions $WUOS$, $WSAT$ and $WTBCU$ form a Voronoi tessellation of the Sierpiński carpet. Let us denote the three regions respectively by M_1 , M_2 and M_3 . If (p_1, p_2) , (q_1, q_2) and (r_1, r_2) are the centroids of these three regions respectively

associated with the probability measure P , we have

$$\begin{aligned}
(p_1, p_2) &= \frac{1}{P(M_1)} \int_{M_1} (x_1, x_2) dP = \frac{1}{P(J_1)} \int_{J_1} (x_1, x_2) dP \\
&= \int (x_1, x_2) dP \circ S_1^{-1} = S_1\left(\frac{1}{2}, \frac{1}{2}\right) = \left(\frac{1}{6}, \frac{1}{6}\right), \\
(q_1, q_2) &= \frac{1}{P(M_2)} \int_{M_2} (x_1, x_2) dP = \frac{1}{P(J_2)} \int_{J_2} (x_1, x_2) dP \\
&= \int (x_1, x_2) dP \circ S_2^{-1} = S_2\left(\frac{1}{2}, \frac{1}{2}\right) = \left(\frac{5}{6}, \frac{1}{6}\right), \\
(r_1, r_2) &= \frac{1}{P(M_3)} \int_{M_3} (x_1, x_2) dP = \frac{1}{P(J_3 \cup J_4)} \int_{J_3 \cup J_4} (x_1, x_2) dP \\
&= \frac{1}{P(J_3 \cup J_4)} \left(\int_{J_3} (x_1, x_2) dP + \int_{J_4} (x_1, x_2) dP \right) \\
&= \frac{1}{P(J_3 \cup J_4)} \left(P(J_3) \int (x_1, x_2) dP \circ S_3^{-1} + P(J_4) \int (x_1, x_2) dP \circ S_4^{-1} \right) \\
&= 2 \left(\frac{1}{4} S_3\left(\frac{1}{2}, \frac{1}{2}\right) + \frac{1}{4} S_4\left(\frac{1}{2}, \frac{1}{2}\right) \right) = \left(\frac{1}{2}, \frac{5}{6}\right).
\end{aligned}$$

Thus, we see that the given set α_3 forms a CVT with three-means. Then use (2), and calculate the corresponding distortion error as

$$\begin{aligned}
&\int \min_{a \in \alpha_3} \|(x_1, x_2) - a\|^2 dP \\
&= \int_{J_1} \|(x_1, x_2) - \left(\frac{1}{6}, \frac{1}{6}\right)\|^2 dP + \int_{J_2} \|(x_1, x_2) - \left(\frac{5}{6}, \frac{1}{6}\right)\|^2 dP \\
&\quad + \int_{J_3 \cup J_4} \|(x_1, x_2) - \left(\frac{1}{2}, \frac{5}{6}\right)\|^2 dP \\
&= \frac{1}{36}V + \frac{1}{36}V + \frac{1}{36} \left(2V + \|S_3\left(\frac{1}{2}, \frac{1}{2}\right) - \left(\frac{1}{2}, \frac{5}{6}\right)\|^2 + \|S_4\left(\frac{1}{2}, \frac{1}{2}\right) - \left(\frac{1}{2}, \frac{5}{6}\right)\|^2 \right) \\
&= \frac{1}{12}.
\end{aligned}$$

Thus, the proof of the lemma is complete. \square

Remark 4.2. The elements in the set $\alpha_3 = \{(\frac{1}{6}, \frac{1}{6}), (\frac{5}{6}, \frac{1}{6}), (\frac{1}{2}, \frac{5}{6})\}$ given by Lemma 4.1 form an isosceles triangle. Due to rotational symmetry there are four such sets giving the same distortion error $\frac{1}{12}$.

Let us now prove the following lemma.

Lemma 4.3. *Let P be a Borel probability measure on \mathbb{R}^2 supported by the Sierpiński carpet as defined before. Then the set $\beta_3 = \{(\frac{5}{6}, \frac{5}{6}), (\frac{13}{90}, \frac{19}{30}), (\frac{19}{30}, \frac{13}{90})\}$ forms a CVT with three-means and the corresponding distortion error is larger than $\frac{1}{12}$.*

Proof. The perpendicular bisectors of the line segments joining each pair of points from the list $\{(\frac{5}{6}, \frac{5}{6}), (\frac{13}{90}, \frac{19}{30}), (\frac{19}{30}, \frac{13}{90})\}$ are SW , OW and TW with equations respectively $x_2 = \frac{979}{405} - \frac{31x_1}{9}$, $x_2 = x_1$ and $x_2 = \frac{979}{1395} - \frac{9x_1}{31}$, and they concur at the point $W(\frac{979}{1800}, \frac{979}{1800})$ as shown in Figure 1 (b). Let (p_1, p_2) , (q_1, q_2) and (r_1, r_2) be the centroids of the three Voronoi regions with centers respectively $P_1(\frac{13}{90}, \frac{19}{30})$, $Q_1(\frac{19}{30}, \frac{13}{90})$ and $R_1(\frac{5}{6}, \frac{5}{6})$. By (3) and (4), with respect to the probability

measure P the centroids of the triangles OBC and OAB are respectively $S_1(\frac{3}{10}, \frac{7}{10}) = (\frac{3}{30}, \frac{7}{30})$ and $S_1(\frac{7}{10}, \frac{3}{10}) = (\frac{7}{30}, \frac{3}{30})$. Therefore, using the definition of centroids, we have

$$\begin{aligned} (p_1, p_2) &= \frac{1}{P(J_3) + P(\triangle OBC)} \left(P(J_3) \int_{J_3} (x_1, x_2) dP + P(\triangle OBC) \int_{\triangle OBC} (x_1, x_2) dP \right) \\ &= \frac{1}{\frac{1}{4} + \frac{1}{8}} \left(\frac{1}{4} \left(\frac{1}{6}, \frac{5}{6} \right) + \frac{1}{8} \left(\frac{3}{30}, \frac{7}{30} \right) \right) = \left(\frac{13}{90}, \frac{19}{30} \right), \\ (q_1, q_2) &= \frac{1}{P(J_2) + P(\triangle OAB)} \left(P(J_2) \int_{J_2} (x_1, x_2) dP + P(\triangle OAB) \int_{\triangle OAB} (x_1, x_2) dP \right) \\ &= \frac{1}{\frac{1}{4} + \frac{1}{8}} \left(\frac{1}{4} \left(\frac{1}{6}, \frac{5}{6} \right) + \frac{1}{8} \left(\frac{7}{30}, \frac{3}{30} \right) \right) = \left(\frac{19}{30}, \frac{13}{90} \right), \\ (r_1, r_2) &= S_4\left(\frac{1}{2}, \frac{1}{2}\right) = \left(\frac{5}{6}, \frac{5}{6}\right). \end{aligned}$$

Thus, we see that the given set β_3 forms a CVT with three-means. Then use (2), and calculate the corresponding distortion error as

$$\begin{aligned} &\int \min_{a \in \beta_3} \|(x_1, x_2) - a\|^2 dP \\ &= \left(\text{distortion error due to the point } \left(\frac{5}{6}, \frac{5}{6}\right) \right) + 2 \left(\text{distortion error due to the point } \left(\frac{13}{90}, \frac{19}{30}\right) \right) \\ &> \frac{1}{36} V + 2 \left(\int_{J_3} \|(x_1, x_2) - \left(\frac{13}{90}, \frac{19}{30}\right)\|^2 dP + \int_{J_{13}} \|(x_1, x_2) - \left(\frac{13}{90}, \frac{19}{30}\right)\|^2 dP \right. \\ &\quad + \int_{J_{113} \cup J_{143}} \|(x_1, x_2) - \left(\frac{13}{90}, \frac{19}{30}\right)\|^2 dP + \int_{J_{1113} \cup J_{1143} \cup J_{1413} \cup J_{1443}} \|(x_1, x_2) - \left(\frac{13}{90}, \frac{19}{30}\right)\|^2 dP \\ &\quad + \int_{J_{11113} \cup J_{11143} \cup J_{11413} \cup J_{11443} \cup J_{14113} \cup J_{14143} \cup J_{14413} \cup J_{14443}} \|(x_1, x_2) - \left(\frac{13}{90}, \frac{19}{30}\right)\|^2 dP \\ &\quad \left. + \int_{J_{111113} \cup J_{111143} \cup J_{111413} \cup J_{111443} \cup J_{141113} \cup J_{141143} \cup J_{141413} \cup J_{141443}} \|(x_1, x_2) - \left(\frac{13}{90}, \frac{19}{30}\right)\|^2 dP \right) \\ &= \frac{1247143}{14929920} = 0.0835331 > 0.0833333 = \frac{1}{12}. \end{aligned}$$

Thus, the proof of the lemma is complete. \square

Remark 4.4. In the CVT β_3 given by Lemma 4.3, one point is the centroid of the child J_4 and the other two points are equidistant from the diagonal passing through the centroid. Due to rotational symmetry of the Sierpiński carpet, there are four such CVTs with three-means in which one point is the centroid of one of the children J_1, J_2, J_3 or J_4 and the other two points are equidistant from the corresponding diagonal, and all have the same distortion error larger than $\frac{1}{12}$.

The following proposition identifies the optimal sets of three-means and associated quantization error.

Proposition 4.5. *Let α_3 be the set given by Lemma 4.1. Then α_3 forms an optimal set of three-means with quantization error $\frac{1}{12}$.*

Proof. The children at each level of the Sierpiński carpet construction are symmetrically distributed over the square, and they each have equal weight with respect to the probability measure P , and so we can say that one point in an optimal set of three-means lies on a line of symmetry of the square and the other two points are equidistant from the line of symmetry.

The square has two different kinds of symmetry: one is a diagonal of the square and one is a perpendicular bisector of the two opposite sides of the square. Comparing Lemma 4.1 and Remark 4.4, we can say that the set α_3 given by Lemma 4.1 forms an optimal set of three-means with quantization error $\frac{1}{12}$. \square

Remark 4.6. Lemma 4.1 and Lemma 4.3 together show that under squared error distortion measure, the centroid condition is not sufficient for optimal quantization for singular continuous probability measures on \mathbb{R}^2 , which is already known for absolutely continuous probability measures (see [DFG]).

5. OPTIMAL SETS OF n MEANS FOR ALL $n \geq 4$

In this section, we discuss about the optimal sets of n -means for all $n \geq 4$. Let α_n be an optimal set of n -means and V_n be the n th quantization error for the probability measure P . From the previous sections, we know that the number of α_2 is two, the number of α_3 is four. The following lemma gives the optimal set of four-means.

Lemma 5.1. *Let P be a Borel probability measure on \mathbb{R}^2 supported by the Sierpiński carpet as defined before. Then the set $\{S_i(\frac{1}{2}, \frac{1}{2}) : 1 \leq i \leq 4\}$ is the only optimal set of four-means with quantization error $\frac{1}{9}V$.*

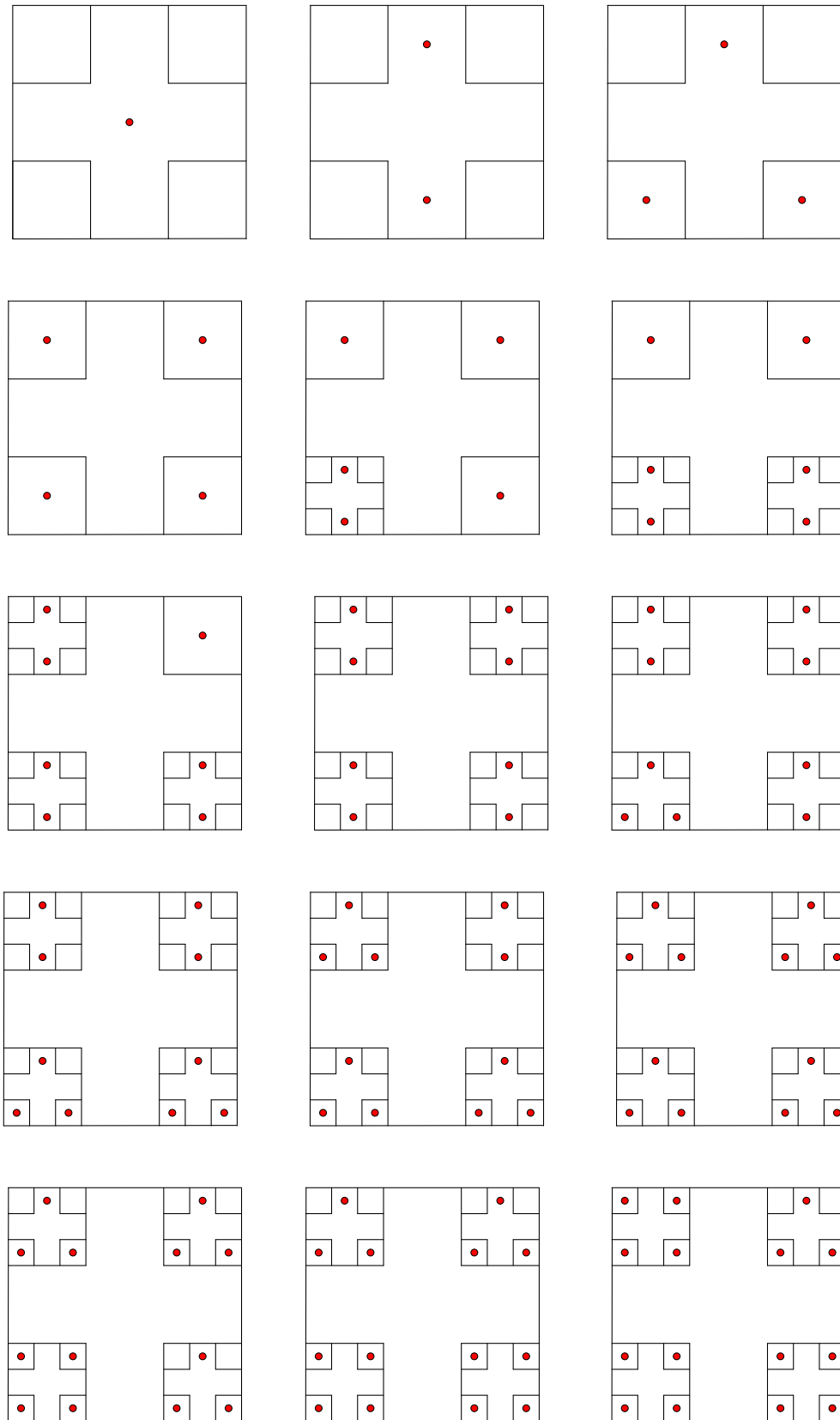
Proof. Let $\alpha = \{a_i : 1 \leq i \leq 4\}$ where $a_i = S_i(\frac{1}{2}, \frac{1}{2})$ for all $1 \leq i \leq 4$. Note that $a_i = E(X : X \in J_i)$. Moreover, J_i are all the children in the first level of the Sierpiński carpet construction, and such that $J_i \subset W(a_i|\alpha)$, i.e., all the points in J_i are closest to $a_i \in \alpha$ for all $1 \leq i \leq 4$. Thus, the set α is the only CVT having the minimum distortion error

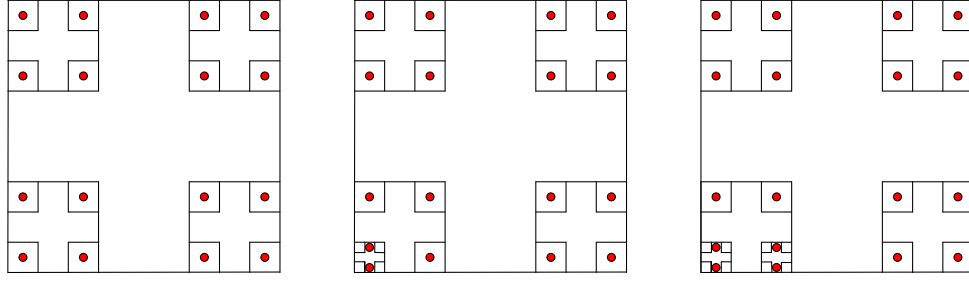
$$\min_{a \in \alpha} \int \|(x_1, x_2) - a\|^2 dP = \sum_{i=1}^4 \int_{J_i} \|(x_1, x_2) - S_i(\frac{1}{2}, \frac{1}{2})\|^2 dP = \frac{1}{9}V.$$

Thus, the lemma is yielded. \square

Lemma 5.2. *Let $n \geq 4$ and α_n be an optimal set of n -means, and let $1 \leq i \leq 4$. Then $\alpha_n \cap J_i \neq \emptyset$, and $\alpha_n \cap (J \setminus J_1 \cup J_2 \cup J_3 \cup J_4)$ is an empty set.*

Proof. If $n = 4$, by Lemma 5.1, we see that $\alpha_n \cap J_i \neq \emptyset$ for all $1 \leq i \leq 4$, in fact, $\alpha_n \cap (J \setminus J_1 \cup J_2 \cup J_3 \cup J_4) = \emptyset$. Let us now prove the lemma for $n = 5$. We first show that $\alpha_5 = \{(\frac{1}{6}, \frac{1}{18}), (\frac{1}{6}, \frac{5}{18}), (\frac{5}{6}, \frac{1}{6}), (\frac{5}{6}, \frac{5}{6})\}$ forms an optimal set of five-means for the probability measure P . As shown in the previous sections, the set $\alpha_2 = \{(\frac{1}{2}, \frac{1}{6}), (\frac{1}{2}, \frac{5}{6})\}$ forms an optimal set of two-means with quantization error $\frac{5}{36}$ and the set $\alpha_1 = \{(\frac{1}{2}, \frac{1}{2})\}$ is an optimal set of one-mean with quantization error $\frac{1}{4}$ for the probability measure P . Since S_i are similarity mappings for all $1 \leq i \leq 4$, the set $S_1(\alpha_2)$, i.e., the set $\{(\frac{1}{6}, \frac{1}{18}), (\frac{1}{6}, \frac{5}{18})\}$ forms an optimal set of two-means for the image measure $P \circ S_1^{-1}$, and the sets $S_i(\alpha_1)$ are the optimal sets of one-mean for the image measures $P \circ S_i^{-1}$ for all $2 \leq i \leq 4$. Recall that $P = \frac{1}{4} \sum_{i=1}^4 P \circ S_i^{-1}$, and proceeding in the same way as in Lemma 4.1, it can be shown that the set $S_1(\alpha_2) \cup S_2(\alpha_1) \cup S_3(\alpha_1) \cup S_4(\alpha_1) = \{(\frac{1}{6}, \frac{1}{18}), (\frac{1}{6}, \frac{5}{18}), (\frac{5}{6}, \frac{1}{6}), (\frac{5}{6}, \frac{5}{6}), (\frac{5}{6}, \frac{5}{6})\}$ forms a CVT with five-means with respect to the probability measure P , and so the set α_5 forms an optimal set of five-means for the probability measure P with quantization error $\frac{5}{36} \frac{1}{36} + \frac{1}{4} \frac{1}{36} + \frac{1}{4} \frac{1}{36} + \frac{1}{4} \frac{1}{36} = \frac{2}{81}$. Note that due to symmetry there are seven more optimal sets of five-means with quantization error $\frac{2}{81}$. From the configuration of α_5 , it is easy to see that the lemma is true for $n = 5$. Using the similar technique inductively, one can prove that the lemma is also true for all $n \geq 6$. Thus, the proof of the lemma is complete. \square

FIGURE 2. Optimal configuration of n points for $1 \leq n \leq 15$.

FIGURE 3. Optimal configuration of n points for $16 \leq n \leq 18$.

The following proposition gives some properties that the optimal sets of n -means and the n th quantization error satisfy for all $n \geq 4$.

Proposition 5.3. *Let $n \geq 4$ be fixed, α_n be an optimal set of n -means, and let $1 \leq i \leq 4$. Set $\beta_i = \alpha_n \cap J_i$ and $n_i = \text{card}(\beta_i)$. Then, $S_i^{-1}(\beta_i)$ is an optimal set of n_i -means. Moreover,*

$$V_n = \frac{1}{36} (V_{n_1} + V_{n_2} + V_{n_3} + V_{n_4}).$$

Proof. By Lemma 5.2, β_i are nonempty for all $1 \leq i \leq 4$. Since α_n does not contain any point from $J \setminus J_1 \cup J_2 \cup J_3 \cup J_4$, we have $\alpha_n = \cup_{i=1}^4 \beta_i$. α_n is an optimal set of n -means, and so

$$V_n = \sum_{i=1}^4 \int_{J_i} \min_{a \in \alpha_n} \|(x_1, x_2) - a\|^2 dP = \sum_{i=1}^4 \int_{J_i} \min_{a \in \beta_i} \|(x_1, x_2) - a\|^2 dP.$$

Now using Lemma 2.1 and the definitions of the mappings S_i , we have

$$(5) \quad V_n = \frac{1}{36} \sum_{i=1}^4 \int \min_{a \in \beta_i} \|(x_1, x_2) - S_i^{-1}(a)\|^2 dP = \frac{1}{36} \sum_{i=1}^4 \int \min_{a \in S_i^{-1}(\beta_i)} \|(x_1, x_2) - a\|^2 dP.$$

If $S_1^{-1}(\beta_1)$ is not an optimal set of n_1 -means, then we can find a set $\gamma_1 \subset \mathbb{R}^2$ with $\text{card}(\gamma_1) = n_1$ such that

$$\int \min_{a \in \gamma_1} \|(x_1, x_2) - a\|^2 dP < \int \min_{a \in S_1^{-1}(\beta_1)} \|(x_1, x_2) - a\|^2 dP.$$

But, then $S_1(\gamma_1) \cup \beta_2 \cup \beta_3 \cup \beta_4$ will be a set of cardinality n , and

$$\begin{aligned} & \int \min\{\|(x_1, x_2) - a\|^2 : a \in S_1(\gamma_1) \cup \beta_2 \cup \beta_3 \cup \beta_4\} dP \\ &= \int_{J_1} \min_{a \in S_1(\gamma_1)} \|(x_1, x_2) - a\|^2 dP + \frac{1}{36} \sum_{i=2}^4 \int \min_{a \in S_i^{-1}(\beta_i)} \|(x_1, x_2) - a\|^2 dP \\ &= \frac{1}{36} \int \min_{a \in S_1(\gamma_1)} \|(x_1, x_2) - S_1^{-1}(a)\|^2 dP + \frac{1}{36} \sum_{i=2}^4 \int \min_{a \in S_i^{-1}(\beta_i)} \|(x_1, x_2) - a\|^2 dP \\ &= \frac{1}{36} \int \min_{a \in \gamma_1} \|(x_1, x_2) - a\|^2 dP + \frac{1}{36} \sum_{i=2}^4 \int \min_{a \in S_i^{-1}(\beta_i)} \|(x_1, x_2) - a\|^2 dP \\ &< \frac{1}{36} \int \min_{a \in S_1^{-1}(\beta_1)} \|(x_1, x_2) - a\|^2 dP + \frac{1}{36} \sum_{i=2}^4 \int \min_{a \in S_i^{-1}(\beta_i)} \|(x_1, x_2) - a\|^2 dP. \end{aligned}$$

Thus by (5), we have $\int \min\{\|(x_1, x_2) - a\|^2 : a \in S_1(\gamma_1) \cup \beta_2 \cup \beta_3 \cup \beta_4\} dP < V_n$, which contradicts the fact that α_n is an optimal set of n -means, and so $S_1^{-1}(\beta_1)$ is an optimal set of n_1 -means.

Similarly, one can show that $S_i^{-1}(\beta_i)$ are optimal sets of n_i -means for all $2 \leq i \leq 4$. Thus, (5) implies $V_n = \frac{1}{36}(V_{n_1} + V_{n_2} + V_{n_3} + V_{n_4})$. This completes the proof of the proposition. \square

By Lemma 2.3 and Lemma 5.1, the sets $\alpha_1 = \{(\frac{1}{2}, \frac{1}{2})\}$ and $\alpha_4 = \{(\frac{1}{6}, \frac{1}{6}), (\frac{5}{6}, \frac{1}{6}), (\frac{1}{6}, \frac{5}{6}), (\frac{5}{6}, \frac{5}{6})\}$ are the only optimal sets of one- and four-means with quantization error $V = \frac{1}{4}$ and $\frac{1}{9}V$, respectively. By Proposition 3.1 and Proposition 4.5, the sets $\alpha_2 = \{(\frac{1}{2}, \frac{1}{6}), (\frac{1}{2}, \frac{5}{6})\}$ and $\alpha_3 = \{(\frac{1}{6}, \frac{1}{6}), (\frac{5}{6}, \frac{1}{6}), (\frac{1}{2}, \frac{5}{6})\}$ are optimal sets of two- and three-means with quantization error $\frac{5}{36}$ and $\frac{1}{12}$, respectively. Also, notice that the sets α_2 and α_3 are not the only optimal sets of two- and three-means; indeed, the total number of optimal sets of two-means is two and the total number of optimal sets of three-means is four. With this, the optimal sets of n -means for all $n \geq 4$, their numbers and the quantization error are given by the following theorem.

Theorem 5.4. *Let P be a Borel probability measure on \mathbb{R}^2 supported by the Sierpiński carpet as defined before. Let $n \in \mathbb{N}$ with $n \geq 4$. For $1 \leq m \leq 3$,*

- (i) *if $n = m4^{\ell(n)}$ for some positive integer $\ell(n)$, then $\alpha_n = \{S_\sigma(\alpha_m) : \sigma \in I^{\ell(n)}\}$ is an optimal set of n -means. The number of such sets is $(2^{m-1})^{4^{\ell(n)}}$ and the corresponding quantization error is given by*

$$V_n = \sum_{\sigma \in I^{\ell(n)}} \int_{J_\sigma} \|(x_1, x_2) - S_\sigma(\alpha_m)\|^2 dP,$$

- (ii) *if $n = m4^{\ell(n)} + k$ where k is a positive integer such that $1 \leq k < 4^{\ell(n)}$ for some positive integer $\ell(n)$, and $t \subset I^{\ell(n)}$ with $\text{card}(t) = k$, then,*

$$\alpha_n(t) = \{S_\sigma(\alpha_m) : \sigma \in I^{\ell(n)} \setminus t\} \cup \{S_\sigma(\alpha_{m+1}) : \sigma \in t\}$$

is an optimal set of n -means. The number of such sets is $(2^{m-1})^{4^{\ell(n)}-k} \cdot 4^{\ell(n)} C_k \cdot 2^{mk}$ if $m = 1, 2$, and $(2^{m-1})^{4^{\ell(n)}-k} \cdot 4^{\ell(n)} C_k$ if $m = 3$; and the corresponding quantization error is given by

$$V_n = \sum_{\sigma \in I^{\ell(n)} \setminus t} \int_{J_\sigma} \|(x_1, x_2) - S_\sigma(\alpha_m)\|^2 dP + \sum_{\sigma \in t} \int_{J_\sigma} \|(x_1, x_2) - S_\sigma(\alpha_{m+1})\|^2 dP,$$

where ${}^u C_v = \binom{u}{v}$, the binomial coefficients.

Proof. We will provide the proof in cases. When $1 \leq n \leq 4$, the assertion follows from Lemma 2.3, Proposition 3.1, Proposition 4.5 and Lemma 5.1 as indicated above. If $n = 4^{\ell(n)}$ for some positive integer $\ell(n)$, then the proof of the statement is similar to Lemma 5.1. If $n \neq 4^{\ell(n)}$ for any positive integer $\ell(n)$, then the proof of the statements given in the theorem can be done by using the similar lines as given in Lemma 5.2 to prove α_5 forms an optimal sets of five-means, and using the properties given by Proposition 5.3. As long as we know the optimal sets of n -means the quantization error can easily be obtained. The number of optimal sets are obtained due to combination of the selections. That is why, we will sketch the main lines of the proof for each of the six cases only.

Case 1. If $n = 4^{\ell(n)}$ for some positive integer $\ell(n)$, then $\alpha_n = \{S_\sigma(\frac{1}{2}, \frac{1}{2}) : \sigma \in I^{\ell(n)}\}$ is the only optimal set of n -means, and the corresponding quantization error is given by

$$\begin{aligned} V_n &= \int \min_{a \in \alpha_n} \|(x_1, x_2) - a\|^2 dP = \sum_{\sigma \in I^{\ell(n)}} \int_{J_\sigma} \|(x_1, x_2) - S_\sigma(\frac{1}{2}, \frac{1}{2})\|^2 dP \\ &= \sum_{\sigma \in I^{\ell(n)}} \frac{1}{4^{\ell(n)}} \int \|(x_1, x_2) - S_\sigma(\frac{1}{2}, \frac{1}{2})\|^2 dP \circ S_\sigma^{-1} = \sum_{\sigma \in I^{\ell(n)}} \frac{1}{4^{\ell(n)}} \frac{1}{9^{\ell(n)}} \int \|(x_1, x_2) - (\frac{1}{2}, \frac{1}{2})\|^2 dP \\ &= \sum_{\sigma \in I^{\ell(n)}} \frac{1}{4^{\ell(n)}} \frac{1}{9^{\ell(n)}} V = \frac{1}{9^{\ell(n)}} V. \end{aligned}$$

Case 2. If $n = 4^{\ell(n)} + k$ where k is a positive integer such that $1 \leq k < 4^{\ell(n)}$ for some positive integer $\ell(n)$, let $t \subset I^{\ell(n)}$ with $\text{card}(t) = k$. Then,

$$\alpha_n(t) = \{S_\sigma(\frac{1}{2}, \frac{1}{2}) : \sigma \in I^{\ell(n)} \setminus t\} \cup \{S_\sigma(\alpha_2) : \sigma \in t\}$$

is an optimal set of n -means. The number of such sets is $4^{\ell(n)} C_k \cdot 2^k$ and

$$\begin{aligned} V_n &= \int \min_{a \in \alpha_n} \|(x_1, x_2) - a\|^2 dP \\ &= \sum_{\sigma \in I^{\ell(n)} \setminus t} \int_{J_\sigma} \|(x_1, x_2) - S_\sigma(\frac{1}{2}, \frac{1}{2})\|^2 dP + \sum_{\sigma \in t} \int_{J_\sigma} \|(x_1, x_2) - S_\sigma(\alpha_2)\|^2 dP \\ &= \frac{1}{36^{\ell(n)}} V \text{card}(I^{\ell(n)} \setminus t) + \sum_{\sigma \in t} \int_{J_\sigma} \|(x_1, x_2) - S_\sigma(\alpha_2)\|^2 dP. \end{aligned}$$

Case 3. If $n = 2 \cdot 4^{\ell(n)}$ for some positive integer $\ell(n)$, then $\alpha_n = \{S_\sigma(\alpha_2) : \sigma \in I^{\ell(n)}\}$ is an optimal set of n -means. The number of such sets is $2^{4^{\ell(n)}}$ and

$$V_n = \sum_{\sigma \in I^{\ell(n)}} \int_{J_\sigma} \|(x_1, x_2) - S_\sigma(\alpha_2)\|^2 dP.$$

Case 4. If $n = 2 \cdot 4^{\ell(n)} + k$ where k is a positive integer such that $1 \leq k < 4^{\ell(n)}$, let $t \subset I^{\ell(n)}$ with $\text{card}(t) = k$. Then,

$$\alpha_n(t) = \{S_\sigma(\alpha_2) : \sigma \in I^{\ell(n)} \setminus t\} \cup \{S_\sigma(\alpha_3) : \sigma \in t\}.$$

is an optimal set of n -means. The number of such sets is $2^{4^{\ell(n)} - k} \cdot 4^{\ell(n)} C_k \cdot 4^k$ and

$$\begin{aligned} V_n &= \int \min_{a \in \alpha_n} \|(x_1, x_2) - a\|^2 dP \\ &= \sum_{\sigma \in I^{\ell(n)} \setminus t} \int_{J_\sigma} \|(x_1, x_2) - S_\sigma(\alpha_2)\|^2 dP + \sum_{\sigma \in t} \int_{J_\sigma} \|(x_1, x_2) - S_\sigma(\alpha_3)\|^2 dP. \end{aligned}$$

Case 5. If $n = 3 \cdot 4^{\ell(n)}$ for some positive integer $\ell(n)$, then $\alpha_n = \{S_\sigma(\alpha_3) : \sigma \in I^{\ell(n)}\}$ is an optimal set of n -means. The number of such sets is $4^{4^{\ell(n)}}$ and

$$V_n = \int \min_{a \in \alpha_n} \|(x_1, x_2) - a\|^2 dP = \sum_{\sigma \in I^{\ell(n)}} \int_{J_\sigma} \|(x_1, x_2) - S_\sigma(\alpha_3)\|^2 dP.$$

Case 6. If $n = 3 \cdot 4^{\ell(n)} + k$ where k is a positive integer such that $1 \leq k < 4^{\ell(n)}$, let $t \subset I^{\ell(n)}$ with $\text{card}(t) = k$. Then,

$$\alpha_n(t) = \{S_\sigma(\alpha_3) : \sigma \in I^{\ell(n)} \setminus t\} \cup \{S_\sigma(\alpha_4) : \sigma \in t\}$$

is an optimal set of n -means. The number of such sets is $4^{4^{\ell(n)}-k} \cdot 4^{\ell(n)} C_k$ and

$$\begin{aligned} V_n &= \int \min_{a \in \alpha_n} \|(x_1, x_2) - a\|^2 dP \\ &= \sum_{\sigma \in I^{\ell(n)} \setminus t} \int_{J_\sigma} \|(x_1, x_2) - S_\sigma(\alpha_3)\|^2 dP + \sum_{\sigma \in t} \int_{J_\sigma} \|(x_1, x_2) - S_\sigma(\alpha_4)\|^2 dP. \end{aligned}$$

This completes all the possible cases; hence, the theorem is proved. \square

Remark 5.5. Previously, Graf and Luschgy determined the optimal sets of n -means and the n th quantization error for a singular continuous probability measure supported by the classical Cantor set C . In this paper, we determined the optimal sets of n -means and the n th quantization error for a singular continuous probability measure supported by a Sierpiński carpet. To the best of our knowledge, the work in this paper is the first advance to investigate the optimal quantizers for a singular continuous probability measure on \mathbb{R}^2 . The technique in this paper can be extended to determine the optimal sets of n -means and the n th quantization error for many other singular continuous probability measures generated by affine transformations in \mathbb{R}^2 .

REFERENCES

- [AW] E.F. Abaya and G.L. Wise, *Some remarks on the existence of optimal quantizers*, Statistics & Probability Letters, Volume 2, Issue 6, December 1984, Pages 349-351.
- [DFG] Q. Du, V. Faber and M. Gunzburger, *Centroidal Voronoi Tessellations: Applications and Algorithms*, SIAM Review, Vol. 41, No. 4 (1999), pp. 637-676.
- [GG] A. Gersho and R.M. Gray, *Vector quantization and signal compression*, Kluwer Academy publishers: Boston, 1992.
- [GKL] R.M. Gray, J.C. Kieffer and Y. Linde, *Locally optimal block quantizer design*, Information and Control, 45 (1980), pp. 178-198.
- [GL] A. György and T. Linder, *On the structure of optimal entropy-constrained scalar quantizers*, IEEE transactions on information theory, vol. 48, no. 2, February 2002.
- [GL1] S. Graf and H. Luschgy, *Foundations of quantization for probability distributions*, Lecture Notes in Mathematics 1730, Springer, Berlin, 2000.
- [GL2] S. Graf and H. Luschgy, *The Quantization of the Cantor Distribution*, Math. Nachr., 183 (1997), 113-133.
- [GN] R.M. Gray and D.L. Neuhoff, *Quantization*, IEEE Transactions on Information Theory, October 1998, Vol. 44 Issue 6, 2325-2383.
- [H] J. Hutchinson, *Fractals and self-similarity*, Indiana Univ. J., 30 (1981), 713-747.
- [Z] R. Zam, *Lattice Coding for Signals and Networks: A Structured Coding Approach to Quantization, Modulation, and Multiuser Information Theory*, Cambridge University Press, 2014.

DEPARTMENT OF MATHEMATICS, 408E24 MINARD HALL, NORTH DAKOTA STATE UNIVERSITY, FARGO, ND 58108-6050, USA.

E-mail address: Dogan.Comez@nds.u.edu

SCHOOL OF MATHEMATICAL AND STATISTICAL SCIENCES, UNIVERSITY OF TEXAS RIO GRANDE VALLEY, 1201 WEST UNIVERSITY DRIVE, EDINBURG, TX 78539-2999, USA.

E-mail address: mrinal.roychowdhury@utrgv.edu